1 Introduction

Motivation:

- LDSC regression is widely used, but it is unclear under which conditions it is consistent
- Dicker / GWASH was developed and conditions given, but those are two complicated and apparently too restrictive.
- The Dicker / GWASH conditions are only sufficient but not necessary.
- Unclear what is the relationship between these estimators, even though in practice they give similar values.
- In this paper we only handle LDSC regression with intercept equal to 1.

Our work:

- The FVE can be defined in three different ways, depending on whether one conditions on $\boldsymbol{X}$ and $\boldsymbol{\beta}$
- Each FVE definition requires a different estimator ???
- All three definitions coincide in the limit under two important conditions on $\boldsymbol{X}$ and $\boldsymbol{\beta}$
- Consistency of the estimators requires those two conditions and a few more on bounded moments of $\boldsymbol{X}$.
- Some estimators may be asymptotically equivalent, others maybe not

Summary of contributions:

- Clear sufficient and necessary conditions for the consistency of LDSC, GWASH and other related estimators
- Clear distinction between FVE definitions depending on conditioning on $\boldsymbol{X}$ and $\boldsymbol{\beta}$
- New estimators for the FVE conditional on $\boldsymbol{\beta}$ or on $\boldsymbol{X}$ and $\boldsymbol{\beta}$???

2 Model and FVE definitions

Suppose a continuous outcome y_i is measured with a panel of m predictors $\vec{\boldsymbol{x}}_i = (x_{i1}, \ldots, x_{im})$ in n independent subjects $i = 1, \ldots, n$. The predictors may be SNP allele counts (taking values 0, 1, 2) or continuous brain measurements. The poly-additive model, called polygenic model in genetics (Fisher, 1918; Lynch and Walsh, 1998), in row or vector form, is

$$y_i = \vec{\boldsymbol{x}}_i \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n, \qquad \text{or} \qquad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where the error terms ε_i are iid with mean 0 and variance σ_ε^2, $\boldsymbol{y} = (y_1, \ldots, y_n)^T$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$, and $\boldsymbol{X}$ is the regression matrix with rows $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_n$ and columns $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$. Typically, model (1) includes other fixed covariates (e.g. age, sex, ethnicity, etc.). For simplicity, we assume that these covariates have been regressed away, so the interpretation of the coefficients is the same as if they had been included in the model. In particular, we assume that the vector $\boldsymbol{y}$ and the columns $\boldsymbol{X}$ have been centered and standardized. Furthermore, it is assumed that the last statement is true both in the population and in the sample. While one can center and standardize only in the sample, when the sample-size is large, as in the typical case we consider, it is reasonable to assume that the centering and standardization carries-out also to the population. We ignore the correlation that is introduced by the the standardization, as it is weak. Let $\boldsymbol{\Sigma} := \mathrm{Cov}(\vec{\boldsymbol{x}}_i)$ denote the $m \times m$ covariance matrix among predictors in the population and $\boldsymbol{R} := \frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}$ is the empirical covariance matrix in the sample. The standardization assumption implies that the diagonal entries of both $\boldsymbol{\Sigma}$ and $\boldsymbol{R}$ are 1. As in Bulik-Sullivan et al. (2015) we assume that $\beta_1, \ldots, \beta_m$ are iid with mean zero and variance h^2/m. Standardization implies that the variance of ε_i is $\sigma_\varepsilon^2 = 1 - h^2$.

In this model the random heritability (or FVE) is $(\vec{\boldsymbol{x}}_i\boldsymbol{\beta})^2$. There are several possible definitions of heritability, which correspond to different ways of taking exception of the random heritability.

- Expected heritability: $h^2 = \mathrm{E}_{\vec{\boldsymbol{x}}_i, \boldsymbol{\beta}}\left[(\vec{\boldsymbol{x}}_i\boldsymbol{\beta})^2\right]$.
- Conditional heritability on the population: $h^2_{pop} := \boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta} = \mathrm{E}_{\vec{\boldsymbol{x}}_i}\left[(\vec{\boldsymbol{x}}_i\boldsymbol{\beta})^2|\boldsymbol{\beta}\right]$.
- Conditional heritability on the sample: $h^2_{sample} := \boldsymbol{\beta}^T\boldsymbol{R}\boldsymbol{\beta} = \hat{\mathrm{E}}_{\vec{\boldsymbol{x}}_i}\left[(\vec{\boldsymbol{x}}_i\boldsymbol{\beta})^2|\boldsymbol{\beta}\right]$, where $\hat{\mathrm{E}}$ is the expectation with respect to the empirical distribution of the sample.

The following theorem shows the relation among the three definitions of heritability.

Theorem 1. *Under model* (1) *the following relations hold,*

1. $h^2_{sample} = h^2$.
2. $\mathrm{E}_{\boldsymbol{\beta}}(h^2_{pop} - h^2)^2 \to 0$ *iff* $\frac{1}{m^2}\mathrm{tr}(\boldsymbol{\Sigma}^2) \to 0$ *(weak correlation) and* $m\mathrm{E}(\beta_i^4) \to 0$ *(distributed effects) (alternatively,* $m(\mathrm{Var}(\beta_i^2) - 2\sigma_\beta^4)$ *).*

Proof of Part 1. We have that

$$\mathrm{E}(y_i^2|\vec{\boldsymbol{x}}_i, \boldsymbol{\beta}) = \mathrm{Var}(y_i|\vec{\boldsymbol{x}}_i, \boldsymbol{\beta}) + [\mathrm{E}(y_i|\vec{\boldsymbol{x}}_i, \boldsymbol{\beta})]^2 = \sigma_\varepsilon^2 + (\vec{\boldsymbol{x}}_i\boldsymbol{\beta})^2 = 1 - h^2 + (\vec{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2.$$

Since the y_i's are standardized in the sample,

$$1 = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(y_i^2|\vec{\boldsymbol{x}}_i, \boldsymbol{\beta}) = 1 - h^2 + \frac{1}{n}\sum_{i=1}^{n}(\vec{\boldsymbol{x}}_i\boldsymbol{\beta})^2 = 1 - h^2 + h^2_{sample}.$$

It follows that $h^2_{sample} = h^2$.

Proof of Part 2. By the law of total expectation, $\mathrm{E}_{\boldsymbol{\beta}}(h^2_{pop}) = h^2$. Therefore, $\mathrm{E}_{\boldsymbol{\beta}}(h^2_{pop} - h^2)^2 = \mathrm{Var}(h^2_{pop}) = \mathrm{Var}(\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta})$. The variance of the quadratic form $\boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta}$ is (Petersen and Pedersen, 2008, Eq.(319)) is

$$2[\mathrm{E}(\beta_i^2)]^2\mathrm{tr}(\boldsymbol{\Sigma}^2) + [(\mathrm{E}(\beta_i^4) - 3[\mathrm{E}(\beta_i^2)]^2)]m = 2h^4\frac{1}{m}\mathrm{tr}(\boldsymbol{\Sigma}^2) + m\mathrm{E}(\beta_i^4) + \frac{h^4}{m}.$$

Therefore, $\mathrm{E}_{\boldsymbol{\beta}}(h^2_{pop} - h^2)^2 \to 0$ iff $\frac{1}{m^2}\mathrm{tr}(\boldsymbol{\Sigma}^2) \to 0$ and $m\mathrm{E}(\beta_i^4) \to 0$. $\square$

Theorem 1 shows that under the model, the definitions of h^2_{sample} and h^2 coincide, and h^2_{pop} is asymptotically equal (in $\mathcal{L}_2$ sense) if and only if two conditions hold. The first condition is $\frac{1}{m^2}\mathrm{tr}(\boldsymbol{\Sigma}^2) \to 0$. Notice that $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ is the Frobenius norm of $\boldsymbol{\Sigma}$. This condition is equivalent to definition of weak correlation that was discussed in Azriel and Schwartzman (2015) in Section 2. It implies that the average over the off-diagonal elements of $\boldsymbol{\Sigma}$ converges to zero; that is, in average $Cov(\vec{\boldsymbol{x}}_{1i}, \vec{\boldsymbol{x}}_{1j})$ is small; for more details see the discussion in Azriel and Schwartzman (2015).

The second condition is $m\mathrm{E}(\beta_i^4) \to 0$. First notice that if the distribution of $\boldsymbol{\beta}$ is normal then $E(\beta_i^4) = 3\frac{h^4}{m^2}$ and the condition is satisfied. To interpret further the second condition, suppose that β_i follows a mixture distribution, where with probability p_m, $\beta_i \sim N(0, \eta_m^2/m)$ and otherwise $\beta_i \sim N(0, \tilde{\eta}_m^2/m)$, where $\tilde{\eta}_m^2 = \frac{h^2 - p_m\eta_m^2}{1-p_m}$. The situation of concentrated effects where a few β's are much larger than the others can be covered in this model when $\eta_m^2 = m\eta_0^2$ and $p_m = C/m$ (for $\eta_0^2 C < h^2$). That is, with a small probability C/m, β_i can be large (i.e., $\beta_i \sim N(0, \eta_0^2)$) and otherwise it is of order $1/m$. In this case, the fourth moment is

$$E(\beta_i^4) = 3\left(C\eta_0^2 + \frac{(h^2 - \eta_0^2)^2}{(1 - C/m)^2 m^2}\right) > 3C\eta_0^2,$$

and the condition in violated. On the other hand, when the effects are uniformly weak, which occurs when η_m^2 and $\tilde{\eta}_m^2$ are bounded, then $mE(\beta_i^4) \to 0$. To sum up, the second condition is satisfied when the β's are all of the same order, and otherwise it may be violated.

3 FVE estimators

Theorem 1 shows that under weak correlation and distributed effects the three definitions of heritability coincide (asymptotically). We now discuss estimation of heritability under these conditions and show that these conditions plays a dominant rule in the consistency of the estimators.

The FVE estimators can be divided into two groups:

The first group contains the estimators GWASH, Dicker's estimator, LDSC ratio (defined below), the u-statistics estimator. The second part of these notes is devoted to show that these estimators are all asymptotically equivalent and they are consistent under weak correlation and distributed effects.

The second group contains the LDSC estimator and variants. They are defined next.

Let $u_j = \frac{1}{\sqrt{n}}\boldsymbol{x}_j^T\boldsymbol{y}$ and $\hat{\ell}_j = \frac{1}{n^2}\boldsymbol{x}_j^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{x}_j$ for $j = 1, ..., m$. They are called the χ^2-statistics and the LD (Linkage Disequilibrium) scores, respectively, in Bulik-Sullivan et al. (2015). Under model (1), we have

$$\begin{aligned}\mathrm{E}(u_j^2|\boldsymbol{X}) &= \mathrm{E}\left\{\frac{1}{n}[\boldsymbol{x}_j^T(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})]^2\Big|\boldsymbol{X}\right\} = \mathrm{E}\left(\frac{1}{n}\boldsymbol{x}_j^T\boldsymbol{X}\boldsymbol{\beta}\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{x}_j + \frac{2}{n}\boldsymbol{x}_j^T\boldsymbol{X}\boldsymbol{\beta}\boldsymbol{x}_j^T\boldsymbol{\varepsilon} + \frac{1}{n}\boldsymbol{x}_j^T\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\boldsymbol{x}_j|\boldsymbol{X}\right)\\ &= \frac{1}{n}\boldsymbol{x}_j^T\boldsymbol{X}\boldsymbol{I}\sigma_\beta^2\boldsymbol{X}^T\boldsymbol{x}_j + \frac{1}{n}\boldsymbol{x}_j^T\boldsymbol{I}\sigma_\varepsilon^2\boldsymbol{x}_j = \frac{h^2}{m}n\hat{\ell}_j + 1 - h^2 = h^2\left(\frac{n}{m}\hat{\ell}_j - 1\right) + 1. \quad (2)\end{aligned}$$

Thus, the conditional expectation of u_j^2 given $\boldsymbol{X}$ is a linear function of $\hat{\ell}_j$ and the slope is h^2. The main estimator we analyze is

$$\hat{h}^2_{LD} := \frac{\bar{\boldsymbol{\ell}}^T\bar{\boldsymbol{u}}}{\bar{\boldsymbol{\ell}}^T\bar{\boldsymbol{\ell}}},$$

where $\bar{\boldsymbol{\ell}} = (\frac{n}{m}\hat{\ell}_1 - 1, \frac{n}{m}\hat{\ell}_2 - 1, \ldots, \frac{n}{m}\hat{\ell}_m - 1)^T$ and $\bar{\boldsymbol{u}} = (u_1^2 - 1, u_2^2 - 1, \ldots, u_m^2 - 1)^T$. One can also consider the mean ratio estimator

$$\hat{h}^2_{LD\,ratio} := \frac{\bar{\boldsymbol{u}}^T\boldsymbol{1}}{\bar{\boldsymbol{\ell}}^T\boldsymbol{1}},$$

where $\boldsymbol{1}$ is the vector of ones. Schwartzman et al. (2019) shows that this estimator is asymptotically equivalent to the estimators of the first group mentioned above. Other variants (THAT SHOULD BE DISCUSSED) are

$$\hat{h}^2_{LD,\boldsymbol{W}} := \frac{\bar{\boldsymbol{\ell}}^T\boldsymbol{W}\bar{\boldsymbol{u}}}{\bar{\boldsymbol{\ell}}^T\boldsymbol{W}\bar{\boldsymbol{\ell}}} \text{ and } \hat{h}^2_{LD,R} := \frac{\tilde{\boldsymbol{\ell}}^T\bar{\boldsymbol{u}}}{\tilde{\boldsymbol{\ell}}^T\tilde{\boldsymbol{\ell}}}.$$

where $\boldsymbol{W}$ is a weighting matrix and $\tilde{\boldsymbol{\ell}}$ is based on LD scored that are coming from a reference dataset and not from the original dataset. The optimal (BLUE) estimator is achieved when $\boldsymbol{W} = [\mathrm{Cov}(\bar{\boldsymbol{u}}|\boldsymbol{X})]^{-1}$.

The estimators:

- Dicker / GWASH
- LDSC regression (LDSC-reg) with intercept = 1
- LDSC ratio (LDSC-ratio)
- U-stat ratio and U-stat regression
- u-score ratio and u-score regression (or chi score maybe)

Motivation for estimators:

- Need to consider conditioning. Show table of the estimators depending on conditioning (on $\boldsymbol{X}$ and $\boldsymbol{\beta}$, on $\boldsymbol{\beta}$ only, or none) and methods (ratio vs regression)

- Some estimators are asymptotically equivalent

- Regression idea allows to modify ratio and address conditioning (simulation with random $\boldsymbol{\beta}$)

Include:

- Definitions

- Asymptotic equivalence

4 Consistency of estimators

The computation of (2) implies that $\mathrm{E}(h^2_{LD}|\boldsymbol{X}) = h^2$. In order to show consistency we need to show that the conditional variance converges to zero. The conditional variance is

$$\mathrm{Var}(h^2_{LD}|\boldsymbol{X}) = \frac{\bar{\boldsymbol{\ell}}^T \mathrm{Cov}(\bar{\boldsymbol{u}}|\mathbf{X})\bar{\boldsymbol{\ell}}}{(\bar{\boldsymbol{\ell}}^T\bar{\boldsymbol{\ell}})^2}$$

Thus, we need to compute $\mathrm{Cov}(\bar{\boldsymbol{u}}|\mathbf{X})$. We shall use the following Lemma.

Lemma 1. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ two fixed symmetric matrices of dimension $d \times d$, and let $\boldsymbol{\xi} = (\xi_1, \dots, \xi_d)^T$ be a vector of iid random variables with mean zero and finite fourth moment. Let $\mathcal{M}_2$ and $\mathcal{M}_4$ denote the second and fourth moment of ξ_i. Then, $Cov(\boldsymbol{\xi}^T\boldsymbol{A}\boldsymbol{\xi}, \boldsymbol{\xi}^T\boldsymbol{B}\boldsymbol{\xi}) = (\mathrm{diag}(\boldsymbol{A}))^T\mathrm{diag}(B)(\mathcal{M}_4 - 3\mathcal{M}_2^2) + 2\mathrm{tr}(\boldsymbol{A}\boldsymbol{B})\mathcal{M}_2^2$.*

We have that

$$\begin{aligned}
&\mathrm{Cov}(u_j^2, u_k^2|\boldsymbol{X}) \\
&= \frac{1}{n^2}\mathrm{Cov}(\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{x}_j\boldsymbol{x}_j^T\boldsymbol{X}\boldsymbol{\beta} + 2\boldsymbol{x}_j^T\boldsymbol{X}\boldsymbol{\beta}\boldsymbol{x}_j^T\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\boldsymbol{x}_j\boldsymbol{x}_j^T\boldsymbol{\varepsilon}, \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{x}_k\boldsymbol{x}_k^T\boldsymbol{X}\boldsymbol{\beta} + 2\boldsymbol{x}_k^T\boldsymbol{X}\boldsymbol{\beta}\boldsymbol{x}_k^T\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\boldsymbol{x}_k\boldsymbol{x}_k^T\boldsymbol{\varepsilon}|\boldsymbol{X}) \\
&\qquad = \frac{1}{n^2}\mathrm{Cov}(\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{x}_j\boldsymbol{x}_j^T\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{x}_k\boldsymbol{x}_k^T\boldsymbol{X}\boldsymbol{\beta}|\boldsymbol{X}) + \frac{1}{n^2}\mathrm{Cov}(\boldsymbol{\varepsilon}^T\boldsymbol{x}_j\boldsymbol{x}_j^T\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}^T\boldsymbol{x}_k\boldsymbol{x}_k^T\boldsymbol{\varepsilon}|\boldsymbol{X}) \\
&\qquad\qquad + \frac{4}{n^2}\mathrm{Cov}(\boldsymbol{x}_j^T\boldsymbol{X}\boldsymbol{\beta}\boldsymbol{x}_j^T\boldsymbol{\varepsilon}, \boldsymbol{x}_k^T\boldsymbol{X}\boldsymbol{\beta}\boldsymbol{x}_k^T\boldsymbol{\varepsilon}|\boldsymbol{X}).
\end{aligned}$$

Let $\boldsymbol{r}_j := \mathbf{X}^T\boldsymbol{x}_j$ and let $\boldsymbol{r}_j^2$ be the element-wise square of $\boldsymbol{r}_j$, for $j = 1, \dots, m$. By Lemma 1,

$$\mathrm{Cov}(\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{x}_j\boldsymbol{x}_j^T\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{x}_k\boldsymbol{x}_k^T\boldsymbol{X}\boldsymbol{\beta}|\boldsymbol{X}) = (\boldsymbol{r}_j^2)^T\boldsymbol{r}_k^2[E(\beta_i^4) - 3h^4/m^2] + 2(\boldsymbol{r}_j^T\boldsymbol{r}_k)^2h^4/m^2.$$

$$\mathrm{Cov}(\boldsymbol{\varepsilon}^T\boldsymbol{x}_j\boldsymbol{x}_j^T\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}^T\boldsymbol{x}_k\boldsymbol{x}_k^T\boldsymbol{\varepsilon}|\boldsymbol{X}) = (\boldsymbol{x}_j^2)^T\boldsymbol{x}_k^2[E(\varepsilon_i^4) - 3(1-h^2)^2] + 2(\boldsymbol{x}_j^T\boldsymbol{x}_k)^2(1-h^2)^2,$$

where $\boldsymbol{x}_j^2$ is the element-wise square of $\boldsymbol{x}_j$, for $j = 1, \dots, m$. The last covariance to compute is

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{x}_j^T \boldsymbol{X}\boldsymbol{\beta}\boldsymbol{x}_j^T\boldsymbol{\varepsilon}, \boldsymbol{x}_k^T \boldsymbol{X}\boldsymbol{\beta}\boldsymbol{x}_k^T\boldsymbol{\varepsilon}|\boldsymbol{X}) = \mathrm{E}(\boldsymbol{\beta}^T \boldsymbol{r}_j \boldsymbol{x}_j^T \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T \boldsymbol{x}_k \boldsymbol{r}_k^T \boldsymbol{\beta}|\boldsymbol{X}) = (1-h^2)\mathrm{E}(\boldsymbol{\beta}^T \boldsymbol{r}_j \boldsymbol{x}_j^T \boldsymbol{x}_k \boldsymbol{r}_k^T \boldsymbol{\beta}|\boldsymbol{X}) \\ = (1-h^2)\frac{h^2}{m}\mathrm{tr}(\boldsymbol{r}_j \boldsymbol{x}_j^T \boldsymbol{x}_k \boldsymbol{r}_k^T) = \frac{(1-h^2)h^2}{m}\boldsymbol{x}_j^T \boldsymbol{x}_k \boldsymbol{r}_k^T \boldsymbol{r}_k.\end{aligned}$$

The conditional variance is

$$\mathrm{Var}(h^2_{LD}|\boldsymbol{X}) = \frac{\sum_{j,k} \bar{\ell}_j \bar{\ell}_k \mathrm{Cov}(u_j^2, u_k^2|\boldsymbol{X})}{(\bar{\boldsymbol{\ell}}^T\bar{\boldsymbol{\ell}})^2}.$$

The term that multiplies $[E(\beta_i^4) - 3h^4/m^2]$ is

$$\frac{1}{n^2(\bar{\boldsymbol{\ell}}^T\bar{\boldsymbol{\ell}})^2}\sum_{j,k}(\boldsymbol{r}_j^2)^T \boldsymbol{r}_k^2 \bar{\ell}_j\bar{\ell}_k = \frac{1}{n^2(\bar{\boldsymbol{\ell}}^T\bar{\boldsymbol{\ell}})^2}\sum_{j,k}(\boldsymbol{r}_j^2)^T \boldsymbol{r}_k^2 \left(\frac{n}{m}\hat{\ell}_j - 1\right)\left(\frac{n}{m}\hat{\ell}_k - 1\right).$$

Recall that $\boldsymbol{r}_j = \boldsymbol{X}^T\boldsymbol{x}_j$; its h-th entry is $r_{j,h} = \boldsymbol{x}_h^T\boldsymbol{x}_j = nR_{j,h}$. Define $\mathcal{R}_{j,h} := (R_{j,h})^2$. Notice that $\hat{\ell}_j = \frac{1}{n^2}\sum_h r_{j,h}^2 = \sum_h \mathcal{R}_{j,h}$. Thus,

$$\begin{aligned}\sum_{j,k}(\boldsymbol{r}_j^2)^T \boldsymbol{r}_k^2 \hat{\ell}_j\hat{\ell}_k = n^4 \sum_{j,k,h,h_1,h_2} \mathcal{R}_{j,h}\mathcal{R}_{k,h}\mathcal{R}_{j,h_1}\mathcal{R}_{k,h_2} = n^4 \sum_h \left(\sum_{j,h_1}\mathcal{R}_{j,h}\mathcal{R}_{j,h_1}\right)\left(\sum_{k,h_2}\mathcal{R}_{k,h}\mathcal{R}_{k,h_2}\right) \\ = n^4\sum_h\left(\sum_{j,h_1}\mathcal{R}_{j,h}\mathcal{R}_{j,h_1}\right)^2 = n^4\sum_h\left(\sum_{h_1}\mathcal{R}^2_{h,h_1}\right)^2;\end{aligned}$$

and,

$$\sum_{j,k}(\boldsymbol{r}_j^2)^T \boldsymbol{r}_k^2 \hat{\ell}_j = n^4 \sum_{j,k,h,h_1} \mathcal{R}_{j,h}\mathcal{R}_{k,h}\mathcal{R}_{j,h_1} = n^4\sum_h\left(\sum_{j,h_1}\mathcal{R}_{j,h}\mathcal{R}_{j,h_1}\right)\left(\sum_k \mathcal{R}_{k,h}\right) = n^4\sum_h\left(\sum_{h_1}\mathcal{R}^2_{h,h_1}\right)\hat{\ell}_h,$$

and,

$$\sum_{j,k}(\boldsymbol{r}_j^2)^T \boldsymbol{r}_k^2 = n^4\sum_{j,k,h}\mathcal{R}_{j,h}\mathcal{R}_{k,h} = n^4\sum_h\sum_{j,h}\mathcal{R}_j\sum_k\mathcal{R}_{k,h} = n^4\sum_h\hat{\ell}_h^2.$$

To sum up the term that multiplies $[E(\beta_i^4) - 3h^4/m^2]$ is

$$\frac{n^2\left(\frac{n}{m}\right)^2\sum_h\left(\sum_{h_1}\mathcal{R}^2_{h,h_1}\right)^2 - 2n^2\left(\frac{n}{m}\right)\sum_h\left(\sum_{h_1}\mathcal{R}^2_{h,h_1}\right)\hat{\ell}_h + n^2\sum_h\hat{\ell}_h^2}{(\bar{\boldsymbol{\ell}}^T\bar{\boldsymbol{\ell}})^2}. \tag{3}$$

Assume that the following quantities are bounded from below and above: $\frac{n}{m}, \sum_{h_1}\mathcal{R}^2_{h,h_1}, \hat{\ell}_h$. Then (3) is bounded above by

$$C_1\left(\sum_h\sum_{h_1}\mathcal{R}^2_{h,h_1} + C_2 m\right) = C_1\left(\|\mathcal{R}^2\|_F^2 + C_2 m\right)$$

It follows that if $\|\mathcal{R}^2\|_F^2 \leq Cm$ (which is a bit stronger than weak correlation) and $mE(\beta_i^4) \to 0$ then this part of the conditional variance goes to zero. On the other hand, I want to show that (3) is bounded from below by Cm. A sufficient condition is that the sum of the first two terms in (3) is positive, but I don't see why this should be true.

- Conditional expectations
- conditional variances
- consistency

Do this for all estimators.

5 Simulations

6 Data example

Ask Anubhav

A comment about weights in LDSC

In the LDSC paper, they computed under the assumptions that both β and ε are normal that

$$Var(\chi_j^2|\ell_j) = (1 + \ell_j N h^2/M)^2.$$

I think that the computation is related to the setting where there are out of sample X's. I got

$$Var(\chi_j^2|X) = (1 - h^2 + \ell_j N h^2/M)^2.$$

More generally, one can consider weighting by the inverse of the conditional covariance matrix. I think that under the normality assumptions we have that

$$Cov(\chi_j^2, \chi_k^2|X) = \big((1 - h^2)S_{j,k} + S_{j,k}^2 N h^2/M\big)^2,$$

where $S = X^T X/n$.

Notes from Feb 2023

Summary: Below the model is defined, the U-statistic is presented, the relation to Dicker's estimate is discussed, and the fourth moment condition is introduced. Specifically, it is (more or less) shown below that:

1. Under Model (4) and Condition (5), the U-statistic estimator of τ_2^2 (denoted by $\hat{\tau}_2^2$) is consistent.
2. The estimator $\hat{\tau}^2 = \frac{\hat{\mu}_1}{\hat{\mu}_2}\hat{\tau}_2^2$ satisfies $\hat{\tau}^2 = \hat{\tau}^2_{Dicker} + O_p(1/n)$ BUT I have a hole in the proof; see the blue comment on page 3.
3. Under a random effects normal mixture model, the estimator is consistent iff $mE(\beta_1^4) \to 0$.

Warning: In several places below the writing is not very formal and also I have a hole in the proof of 2.

Introduction: Suppose a continuous outcome y_i is measured with a panel of m predictors $\vec{\boldsymbol{x}}_i = (x_{i1}, \dots, x_{im})$ in n independent subjects $i = 1, \dots, n$. The predictors may be SNP allele counts (taking values 0, 1, 2) or continuous brain measurements. The poly-additive model is

$$y_i = \vec{\boldsymbol{x}}_i \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \dots, n, \qquad \text{or} \qquad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{4}$$

where the error terms ε_i are iid with mean 0 and variance σ^2, $\boldsymbol{y} = (y_1, \dots, y_n)^T$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$, and $\boldsymbol{X}$ is the regression matrix with rows $\vec{\boldsymbol{x}}_1, \dots, \vec{\boldsymbol{x}}_n$ and columns $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$. It is assumed that σ^2 does not depend on n or p. Let $\boldsymbol{\Sigma} := E(\vec{\boldsymbol{x}}_1^T \vec{\boldsymbol{x}}_1)$ denote the covariance matrix of $\vec{\boldsymbol{x}}_i$.

Estimates for $\mathrm{tr}(\boldsymbol{\Sigma})/m$ and $\mathrm{tr}(\boldsymbol{\Sigma}^2)/m$: Let $\hat{\mu}_1 = \frac{1}{mn}\sum_{i=1}^n \vec{\boldsymbol{x}}_i \vec{\boldsymbol{x}}_i^T$ and $\hat{\mu}_2 = \frac{2}{mn(n-1)}\sum_{i_1<i_2}(\vec{\boldsymbol{x}}_{i_1}\vec{\boldsymbol{x}}_{i_2}^T)^2$, be U-statistic estimates of $\mathrm{tr}(\boldsymbol{\Sigma})/m$ and $\mathrm{tr}(\boldsymbol{\Sigma}^2)/m$. It is easy to see that $\hat{\mu}_1$ is unbiased; also $\hat{\mu}_2$ is unbiased because

$$E\left[(\vec{\boldsymbol{x}}_{i_1}\vec{\boldsymbol{x}}_{i_2}^T)^2\right] = E\left[\vec{\boldsymbol{x}}_{i_1}\vec{\boldsymbol{x}}_{i_2}^T\vec{\boldsymbol{x}}_{i_2}\vec{\boldsymbol{x}}_{i_1}^T\right] = E\left[\mathrm{tr}(\vec{\boldsymbol{x}}_{i_1}^T\vec{\boldsymbol{x}}_{i_1}\vec{\boldsymbol{x}}_{i_2}^T\vec{\boldsymbol{x}}_{i_2})\right] = \mathrm{tr}E\left[\vec{\boldsymbol{x}}_{i_1}^T\vec{\boldsymbol{x}}_{i_1}\right]E\left[\vec{\boldsymbol{x}}_{i_2}^T\vec{\boldsymbol{x}}_{i_2}\right] = \mathrm{tr}(\boldsymbol{\Sigma}^2).$$

By Thm. 12.3 of van der Vaart (2000),

$$Var\left(\hat{\mu}_2\right) = \frac{1}{m^2}\left\{\frac{4(n-2)}{n(n-1)}\zeta_1 + \frac{2}{n(n-1)}\zeta_2\right\},$$

where $\zeta_1 = Cov((\vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_2^T)^2, (\vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_3^T)^2)$ and $\zeta_2 = Var\left((\vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_2^T)^2\right)$. We have that

$$\zeta_1 = Cov((\vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_2^T)^2, (\vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_3^T)^2) = E(\vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_2^T\vec{\boldsymbol{x}}_2\vec{\boldsymbol{x}}_1^T\vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_3^T\vec{\boldsymbol{x}}_3\vec{\boldsymbol{x}}_1^T) - \{\mathrm{tr}(\boldsymbol{\Sigma}^2)\}^2 = E(\vec{\boldsymbol{x}}_1\boldsymbol{\Sigma}\vec{\boldsymbol{x}}_1^T)^2 - \{\mathrm{tr}(\boldsymbol{\Sigma}^2)\}^2,$$

and,

$$\zeta_2 = Var\left((\vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_2^T)^2\right) = E\left((\vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_2^T)^4\right) - \{\mathrm{tr}(\boldsymbol{\Sigma}^2)\}^2.$$

A comment about the relation between $\hat{\mu}_2$ and Dicker's estimate for $\mathrm{tr}(\boldsymbol{\Sigma}^2)/m$: Dicker's estimate for $\mathrm{tr}(\boldsymbol{\Sigma}^2)/m$ is

$$\hat{m}_2 = \frac{1}{m}\left\{\mathrm{tr}\left(\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}\right)^2\right\} - \frac{1}{mn}\left\{\mathrm{tr}\left(\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X}\right)\right\}^2 = \frac{1}{mn^2}\left[\mathrm{tr}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^2 - \frac{1}{n}\left\{\mathrm{tr}\left(\boldsymbol{X}^T\boldsymbol{X}\right)\right\}^2\right].$$

We have that

$$\mathrm{tr}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^2 = \mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^T\right)^2 = \sum_{i_1,i_2}(\vec{\boldsymbol{x}}_{i_1}\vec{\boldsymbol{x}}_{i_2}^T)^2,$$

and,

$$\mathrm{tr}\left(\boldsymbol{X}^T\boldsymbol{X}\right) = \mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^T\right) = \sum_i \vec{\boldsymbol{x}}_i\vec{\boldsymbol{x}}_i^T.$$

Therefore,

$$\frac{n}{n-1}\hat{\tau}_2^2 - \hat{m}_2 = \frac{1}{mn}\left\{\frac{1}{n}\sum_i(\vec{\boldsymbol{x}}_i\vec{\boldsymbol{x}}_i^T)^2 - \left(\frac{1}{n}\sum_i\vec{\boldsymbol{x}}_i\vec{\boldsymbol{x}}_i^T\right)^2\right\}.$$

The term in the curly brackets is the empirical variance of $\vec{\boldsymbol{x}}_i\vec{\boldsymbol{x}}_i^T$; we have that $Var(\vec{\boldsymbol{x}}_i\vec{\boldsymbol{x}}_i^T) = \mathrm{tr}(\boldsymbol{\Sigma}) = m$. Therefore,

$$\frac{n}{n-1}\hat{\tau}_2^2 - \hat{m}_2 \approx \frac{m}{mn} = \frac{1}{n},$$

which implies that $\hat{\tau}_2^2 - \hat{m}_2 = O(1/n)$.

A U-statistic estimate of τ_2^2: Let $\vec{\boldsymbol{w}}_i = \vec{\boldsymbol{x}}_i y_i$. Define $\hat{\tau}_2^2 = \frac{2}{n(n-1)}\sum_{i_1<i_2}\vec{\boldsymbol{w}}_{i_1}\vec{\boldsymbol{w}}_{i_2}^T$. We have that

$$E\left[\vec{\boldsymbol{w}}_1\vec{\boldsymbol{w}}_2^T\right] = E\left[(\boldsymbol{\beta}^T\vec{\boldsymbol{x}}_1^T + \varepsilon_1)\vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_2^T(\vec{\boldsymbol{x}}_2\boldsymbol{\beta} + \varepsilon_2)\right] = \boldsymbol{\beta}^T\boldsymbol{\Sigma}^2\boldsymbol{\beta},$$

and therefore, $\hat{\tau}_2^2$ is unbiased for τ_2^2. Its variance is

$$Var\left(\hat{\tau}_2^2\right) = \left\{\frac{4(n-2)}{n(n-1)}\eta_1 + \frac{2}{n(n-1)}\eta_2\right\},$$

where

$$\begin{aligned}\eta_1 = Cov(\vec{\boldsymbol{w}}_1\vec{\boldsymbol{w}}_2^T, \vec{\boldsymbol{w}}_1\vec{\boldsymbol{w}}_3^T)\\ = E\left[(\boldsymbol{\beta}^T\vec{\boldsymbol{x}}_2^T + \varepsilon_2)\vec{\boldsymbol{x}}_2\vec{\boldsymbol{x}}_1^T(\vec{\boldsymbol{x}}_1\boldsymbol{\beta} + \varepsilon_1)(\boldsymbol{\beta}^T\vec{\boldsymbol{x}}_1^T + \varepsilon_1)\vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_3^T(\vec{\boldsymbol{x}}_3\boldsymbol{\beta} + \varepsilon_3)\right] - \left(\boldsymbol{\beta}^T\boldsymbol{\Sigma}^2\boldsymbol{\beta}\right)^2\\ = \boldsymbol{\beta}^T\boldsymbol{\Sigma}E\left[\vec{\boldsymbol{x}}_1^T\vec{\boldsymbol{x}}_1(\vec{\boldsymbol{x}}_1\boldsymbol{\beta} + \varepsilon_1)^2\right]\boldsymbol{\Sigma}\boldsymbol{\beta} - \left(\boldsymbol{\beta}^T\boldsymbol{\Sigma}^2\boldsymbol{\beta}\right)^2\\ = \boldsymbol{\beta}^T\boldsymbol{\Sigma}\left\{E\left[\vec{\boldsymbol{x}}_1^T\vec{\boldsymbol{x}}_1(\vec{\boldsymbol{x}}_1\boldsymbol{\beta})^2\right] + \sigma^2\boldsymbol{\Sigma}\right\}\boldsymbol{\Sigma}\boldsymbol{\beta} - \left(\boldsymbol{\beta}^T\boldsymbol{\Sigma}^2\boldsymbol{\beta}\right)^2\end{aligned}$$

and

$$\begin{aligned}\eta_2 = Var(\vec{\boldsymbol{w}}_1\vec{\boldsymbol{w}}_2^T) = E\left[\left\{(\boldsymbol{\beta}^T\vec{\boldsymbol{x}}_2^T + \varepsilon_2)\vec{\boldsymbol{x}}_2\vec{\boldsymbol{x}}_1^T(\vec{\boldsymbol{x}}_1\boldsymbol{\beta} + \varepsilon_1)\right\}^2\right] - \left(\boldsymbol{\beta}^T\boldsymbol{\Sigma}^2\boldsymbol{\beta}\right)^2 =\\ E\left[(\vec{\boldsymbol{x}}_2\vec{\boldsymbol{x}}_1^T)^2\left\{(\vec{\boldsymbol{x}}_1\boldsymbol{\beta})^2(\vec{\boldsymbol{x}}_2\boldsymbol{\beta})^2 + \sigma^2((\vec{\boldsymbol{x}}_1\boldsymbol{\beta})^2 + (\vec{\boldsymbol{x}}_2\boldsymbol{\beta})^2) + \sigma^4\right\}\right] - \left(\boldsymbol{\beta}^T\boldsymbol{\Sigma}^2\boldsymbol{\beta}\right)^2.\end{aligned}$$

Sufficient conditions for consistency: Assume Model (4) and suppose that there exists a constant C such that

$$\vec{\boldsymbol{x}}_i\vec{\boldsymbol{x}}_i^T \le Cm \text{ for all } i;\ (\vec{\boldsymbol{x}}_i\boldsymbol{\beta})^2 \le C \text{ for all } i, \text{ and } \lambda_{max}(\boldsymbol{\Sigma}) \le C. \tag{5}$$

Then $\hat{\tau}_2^2 - \tau_2^2 \xrightarrow{p} 0$ when $n, m \to \infty$ where $n/m \to \alpha \in [0, \infty)$.

Proof It is enough to show that

$$\frac{\zeta_1}{m^2 n} \to 0, \ \frac{\zeta_2}{m^2 n^2} \to 0, \ \frac{\eta_1}{n} \to 0, \text{ and } \frac{\eta_2}{n^2} \to 0.$$

Indeed,

$$\frac{\zeta_1}{m^2 n} \leq \frac{E(\vec{\boldsymbol{x}}_1 \boldsymbol{\Sigma} \vec{\boldsymbol{x}}_1^T)^2}{m^2 n} \leq \lambda_{max}^2 \frac{E(\vec{\boldsymbol{x}}_1 \vec{\boldsymbol{x}}_1^T)^2}{m^2 n} \leq C\lambda_{max}^2 \frac{E(\vec{\boldsymbol{x}}_1 \vec{\boldsymbol{x}}_1^T)}{mn} = C\lambda_{max}^2 \frac{\text{tr}(\boldsymbol{\Sigma})}{mn} \leq C\lambda_{max}^3 \frac{1}{n} \to 0,$$

where the first inequality follows from the definition of ζ_1, the second from the Rayleigh quotient inequality, the third from the assumption that $\vec{\boldsymbol{x}}_i \vec{\boldsymbol{x}}_i^T \leq Cm$ and the last inequality follows from $\frac{\text{tr}(\boldsymbol{\Sigma})}{m} \leq \lambda_{max}$. Now,

$$\frac{\zeta_2}{m^2 n^2} \leq \frac{E\left((\vec{\boldsymbol{x}}_1 \vec{\boldsymbol{x}}_2^T)^4\right)}{m^2 n^2} \leq \frac{E\left((\vec{\boldsymbol{x}}_1 \vec{\boldsymbol{x}}_2^T)^2\right)}{n^2} = \frac{E(\vec{\boldsymbol{x}}_1 \boldsymbol{\Sigma} \vec{\boldsymbol{x}}_1^T)}{n^2} \leq \lambda_{max} \frac{E(\vec{\boldsymbol{x}}_1 \vec{\boldsymbol{x}}_1^T)}{n^2} \leq C\lambda_{max} \frac{m}{n^2} \to 0,$$

where the first inequality follows from the definition of ζ_2, and the second Cauchy-Schwartz and from the assumption that $\vec{\boldsymbol{x}}_i \vec{\boldsymbol{x}}_i^T \leq Cm$. Also,

$$\begin{aligned}\frac{\eta_1}{n} &\leq \frac{\boldsymbol{\beta}^T \boldsymbol{\Sigma} \left\{ E\left[\vec{\boldsymbol{x}}_1^T \vec{\boldsymbol{x}}_1 (\vec{\boldsymbol{x}}_1 \boldsymbol{\beta})^2 \right] + \sigma^2 \boldsymbol{\Sigma} \right\} \boldsymbol{\Sigma} \boldsymbol{\beta}}{n} \leq C \frac{\boldsymbol{\beta}^T \boldsymbol{\Sigma} E\left[\vec{\boldsymbol{x}}_1^T \vec{\boldsymbol{x}}_1 \right] \boldsymbol{\Sigma} \boldsymbol{\beta}}{n} + \sigma^2 \frac{\boldsymbol{\beta}^T \boldsymbol{\Sigma}^3 \boldsymbol{\beta}}{n} \\ &= \frac{\boldsymbol{\beta}^T \boldsymbol{\Sigma}^3 \boldsymbol{\beta} (C + \sigma^2)}{n} \leq \lambda_{max}^2 \frac{\boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} (C + \sigma^2)}{n} = \lambda_{max}^2 \frac{E((\vec{\boldsymbol{x}}_1 \boldsymbol{\beta})^2)(C + \sigma^2)}{n} \leq \lambda_{max}^2 \frac{C(C + \sigma^2)}{n} \to 0.\end{aligned}$$

Finally,

$$\begin{aligned}\frac{\eta_2}{n^2} &\leq \frac{E\left[(\vec{\boldsymbol{x}}_2 \vec{\boldsymbol{x}}_1^T)^2 \left\{ (\vec{\boldsymbol{x}}_1 \boldsymbol{\beta})^2 (\vec{\boldsymbol{x}}_2 \boldsymbol{\beta})^2 + \sigma^2 ((\vec{\boldsymbol{x}}_1 \boldsymbol{\beta})^2 + (\vec{\boldsymbol{x}}_2 \boldsymbol{\beta})^2) + \sigma^4 \right\} \right]}{n^2} \leq (C^2 + \sigma^2)^2 \frac{E\left[(\vec{\boldsymbol{x}}_2 \vec{\boldsymbol{x}}_1^T)^2 \right]}{n^2} \\ &= (C^2 + \sigma^2)^2 \frac{\text{tr}(\boldsymbol{\Sigma}^2)}{n^2} \leq (C^2 + \sigma^2)^2 \frac{\lambda_{max}^2 m}{n^2} \to 0.\end{aligned}$$

The relation between The U-statistic estimator and Dicker's The U-statistic estimator is

$$\hat{\tau}^2 = \frac{\hat{\mu}_1}{\hat{\mu}_2} \hat{\tau}_2^2 = \frac{\hat{\mu}_1}{\hat{\mu}_2} \frac{1}{n(n-1)} \sum_{i_1 \neq i_2} \vec{\boldsymbol{w}}_{i_1} \vec{\boldsymbol{w}}_{i_2}^T,$$

and Dicker's estimate is

$$\hat{\tau}_{Dicker}^2 = \frac{\hat{m}_1}{\hat{m}_2} \frac{1}{n(n+1)} \left(\|\boldsymbol{X}^T \boldsymbol{y}\|^2 - m\hat{m}_1 \|\boldsymbol{y}\|^2 \right).$$

Notice that $\hat{\mu}_1 = \hat{m}_1$ and we saw that $\hat{\tau}_2^2 = \hat{m}_2 + O_p(1/n)$. Also,

$$\|\boldsymbol{X}^T \boldsymbol{y}\|^2 = \sum_{j=1}^m (\boldsymbol{x}_j^T \boldsymbol{y})^2 = \sum_{j=1}^m \left(\sum_{i=1}^n \boldsymbol{X}_{ij} y_i \right)^2 = \sum_{j=1}^m \sum_{i_1, i_2} \boldsymbol{X}_{i_1 j} \boldsymbol{X}_{i_2 j} y_{i_1} y_{i_2} = \sum_{i_1, i_2} \vec{\boldsymbol{w}}_{i_1} \vec{\boldsymbol{w}}_{i_2}^T.$$

Thus, we have that

$$\begin{aligned}\frac{(n+1)\hat{m}_2}{(n-1)\hat{\mu}_2} \hat{\tau}_{Dicker}^2 - \hat{\tau}^2 &= \frac{\hat{\mu}_1}{\hat{\mu}_2} \frac{\|\boldsymbol{X}^T \boldsymbol{y}\|^2 - m\hat{m}_1 \|\boldsymbol{y}\|^2 - \sum_{i_1 \neq i_2} \vec{\boldsymbol{w}}_{i_1} \vec{\boldsymbol{w}}_{i_2}^T}{n(n-1)} = \frac{\hat{\mu}_1}{\hat{\mu}_2} \frac{\sum_{i=1}^n \vec{\boldsymbol{w}}_i \vec{\boldsymbol{w}}_i^T - m\hat{m}_1 \|\boldsymbol{y}\|^2}{n(n-1)} \\ &= \frac{\hat{\mu}_1}{\hat{\mu}_2} \frac{1}{n-1} \left\{ \frac{1}{n} \sum_{i=1}^n y_i^2 \vec{\boldsymbol{x}}_i \vec{\boldsymbol{x}}_i^T - \frac{1}{n} \sum_{i=1}^n \vec{\boldsymbol{x}}_i \vec{\boldsymbol{x}}_i^T \frac{1}{n} \sum_{i=1}^n y_i^2 \right\}.\end{aligned}$$

The term in the curly brackets is the empirical covariance of y_1^2 and $\vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_1^T$, and it converges to $Cov((\vec{\boldsymbol{x}}_1\boldsymbol{\beta})^2, \vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_1^T)$.
NEED TO SHOW: $Cov((\vec{\boldsymbol{x}}_1\boldsymbol{\beta})^2, \vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_1^T)$ does not depend on m! (or at least is bounded)

An attempt Let $\boldsymbol{\Gamma}$ denote the $m\times m$ matrix, where $\Gamma_{ij} = Cov(x_{1i}x_{1j}, \|\vec{\boldsymbol{x}}_1\|^2)$. Then, $Cov((\vec{\boldsymbol{x}}_1\boldsymbol{\beta})^2, \vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_1^T) = \boldsymbol{\beta}^T\boldsymbol{\Gamma}\boldsymbol{\beta}$. Thus, a sufficient condition for $Cov((\vec{\boldsymbol{x}}_1\boldsymbol{\beta})^2, \vec{\boldsymbol{x}}_1\vec{\boldsymbol{x}}_1^T)$ being bounded is that $\lambda_{max}(\boldsymbol{\Gamma})$ is bounded.
An example: Suppose that $\vec{\boldsymbol{x}} \sim N(\vec{\boldsymbol{0}}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ is a correlation matrix. Then,

$$Cov(x_{1i}x_{1j}, x_{1k}^2) = E(x_{1i}x_{1j}x_{1k}^2) - E(x_{1i}x_{1j}) = 2E(x_{1i}x_{1k})E(x_{1j}x_{1k}) = 2\Sigma_{ik}\Sigma_{j,k},$$

where the second equality follows from a formula in Wikipedia. Since $\boldsymbol{\Gamma}_{ij} = \sum_k Cov(x_{1i}x_{1j}, x_{1k}^2)$, it follows that $\boldsymbol{\Gamma} = 2\boldsymbol{\Sigma}^2$. Therefore, $\lambda_{max}(\boldsymbol{\Gamma}) = \lambda_{max}(\boldsymbol{\Sigma})^2$.

It follows that

$$\hat{\tau}^2_{Dicker} - \hat{\tau}^2 = O_p(1/n).$$

The fourth moment condition For the above estimates to be consistent we need the following condition to hold

$$\tau^2 - \frac{\mu_1}{\mu_2}\tau_2^2 \to 0, \tag{6}$$

as $m \to \infty$; recall that $\tau^2 = \boldsymbol{\beta}^T\boldsymbol{\Sigma}\boldsymbol{\beta}$, $\tau_2^2 = \boldsymbol{\beta}^T\boldsymbol{\Sigma}^2\boldsymbol{\beta}$, $\mu_1 = \mathrm{tr}(\boldsymbol{\Sigma})/m$ and $\mu_2 = \mathrm{tr}(\boldsymbol{\Sigma}^2)/m$. In order to study Condition (6) we model $\boldsymbol{\beta}$ as iid with distribution that is a mixture (possibly infinite) of normals with mean zero; specifically,

$$P(\beta_1 \le t) = \sum_k \pi_k \Phi(t/\sigma_k),$$

where $\pi_k > 0$, $\sum_k \pi_k = 1$, $\sigma_k > 0$ and $\Phi(\cdot)$ is the CDF of a standard normal. We assume that $\tilde{C} = E(\tau^2) = \mathrm{tr}(\boldsymbol{\Sigma})\sum_k \pi_k\sigma_k^2 = m\mu_1\sum_k \pi_k\sigma_k^2$; thus we assume that $\sum_k \pi_k\sigma_k^2 = C/m$, where $C = \tilde{C}/\mu_1$.

We have that $E(\tau^2) = \mathrm{tr}(\boldsymbol{\Sigma}^2)\sum_k \pi_k\sigma_k^2 = m\mu_2\sum_k \pi_k\sigma_k^2 = \mu_2 C$. Therefore,

$$E(\tau^2 - \frac{\mu_1}{\mu_2}\tau_2^2) = C\mu_1 - \frac{\mu_1}{\mu_2}C\mu_2 = 0.$$

Thus, in order for (6) to hold with high-probability we need $Var(\tau^2 - \frac{\mu_1}{\mu_2}\tau_2^2) \to 0$. Now, let K denote the random component of the mixture of $\boldsymbol{\beta}$. The law of total variance implies that

$$\begin{aligned} Var(\tau^2) = E(Var(\tau^2|K)) + Var(E(\tau^2|K)) &= 2\mathrm{tr}(\boldsymbol{\Sigma}^2)\sum_k \pi_k\sigma_k^4 + Var(\mathrm{tr}(\boldsymbol{\Sigma})\sigma_K^2) \\ &= 2m\mu_2\sum_k \pi_k\sigma_k^4 + m^2\mu_1^2\left[\sum_k \sigma_k^4\pi_k - \left(\sum_k \pi_k\sigma_k^2\right)^2\right]. \end{aligned}$$

Similarly,

$$Var(\tau_2^2) = 2m\mu_4\sum_k \pi_k\sigma_k^4 + m^2\mu_2^2\left[\sum_k \sigma_k^4\pi_k - \left(\sum_k \pi_k\sigma_k^2\right)^2\right],$$

where $\mu_4 = \mathrm{tr}(\mathbf{\Sigma}^4)/m$. Finally,

$$Cov(\tau^2, \tau_2^2) = E(Cov(\tau^2, \tau_2^2|K)) + Cov(E(\tau^2|K), E(\tau_2^2|K)) = 2\mathrm{tr}(\mathbf{\Sigma}^3)E(\sigma_j^4) + Cov(\mu_1 m\sigma_K^2, \mu_2 m\sigma_K^2) =$$
$$2m\mu_3 \sum_k \pi_k \sigma_k^4 + m^2\mu_1\mu_2 \left[\sum_k \sigma_k^4 \pi_k - \left(\sum_k \pi_k \sigma_k^2\right)^2\right].$$

It follows that

$$Var\left(\tau^2 - \frac{\mu_1}{\mu_2}\tau_2^2\right) = Var(\tau^2) + \frac{\mu_1^2}{\mu_2^2}Var(\tau_2^2) - 2\frac{\mu_1}{\mu_2}Cov(\tau^2, \tau_2^2)$$
$$= 2m\left(\mu_2 + \frac{\mu_1^2}{\mu_2^2}\mu_4 - 2\frac{\mu_1}{\mu_2}\mu_3\right)\sum_k \pi_k \sigma_k^4.$$

Notice that the term in the square brackets cancels out; notice also that $E(\beta_1^4) = 2\sum_k \pi_k \sigma_k^4$. Therefore, assuming that $\mu_2 + \frac{\mu_1^2}{\mu_2^2}\mu_4 - 2\frac{\mu_1}{\mu_2}\mu_3$ is a positive constant, we have that

$$Var\left(\tau^2 - \frac{\mu_1}{\mu_2}\tau_2^2\right) \Longleftrightarrow mE(\beta_1^4) \to 0. \tag{7}$$

A comment about the term $\mu_2 + \frac{\mu_1^2}{\mu_2^2}\mu_4 - 2\frac{\mu_1}{\mu_2}\mu_3$: If V is a random variable that is uniform on the eigenvalues $\lambda_1, \ldots, \lambda_m$, then its moments are $\mu_1, \mu_2, \mu_3, \mu_4$ and

$$E\left(V - \frac{\mu_1}{\mu_2}V^2\right)^2 = \mu_2 + \frac{\mu_1^2}{\mu_2^2}\mu_4 - 2\frac{\mu_1}{\mu_2}\mu_3.$$

Thus, this term is zero iff $V - \frac{\mu_1}{\mu_2}V^2 \equiv 0$ iff V is a constant (that is, iff all eigenvalues are the same).

References

Azriel, D. and A. Schwartzman (2015). The empirical distribution of a large number of correlated normal variables. *Journal of the American Statistical Association 110*(511), 1217–1228.

Bulik-Sullivan, B. K., P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, S. W. G. of the Psychiatric Genomics Consortium, N. Patterson, M. J. Daly, A. L. Price, and B. M. Neale (2015). Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics 47*(3), 291–295.

Bulik-Sullivan, B. K., P. R. Loh, H. K. Finucane, S. Ripke, J. Yang, C. Schizophrenia Working Group of the Psychiatric Genomics, N. Patterson, M. J. Daly, A. L. Price, and B. M. Neale (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics 47*(3), 291.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh 52*, 399–433.

Lynch, M. and B. Walsh (1998). *Genetics and analysis of quantitative traits. Vol. 1.* Sinauer Sunderland, MA.

Petersen, K. B. and M. S. Pedersen (2008). The matrix cookbook. *Technical University of Denmark 7*(15), 510.

Schwartzman, A., A. J. Schork, R. Zablocki, and W. K. Thompson (2019). A simple, consistent estimator of SNP heritability from genome-wide association studies. *The Annals of Applied Statistics 13*(4), 2509–2538.